

Universitat Politècnica de Catalunya

FACULTAT D'INFORMÀTICA DE BARCELONA

ESPECIALITAT DE COMPUTACIÓ

Graph Neural Networks (GNN): Aplicacions en els
sistemes financers

TREBALL DE FI DE GRAU
ENTREGA 4: DOCUMENT FINAL

MIQUEL MUÑOZ GARCÍA-RAMOS

DIRECTOR: SERGI ABADAL
CODIRECTOR: AXEL WASSINGTON
TUTOR GEP: ANTONIO CAÑABATE

20 de març de 2024



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona



Índex

1	Introducció i contextualització	1
1.1	Definició de conceptes	2
1.1.1	Intel·ligència artificial i aprenentatge automàtic	2
1.1.2	<i>Graph Neural Network</i> (GNN)	2
1.2	Identificació del problema i estat de l'art	3
1.3	Agents implicats	7
2	Justificació	7
3	Abast	9
3.1	Objectius i subobjectius	9
3.2	Requeriments funcionals i no funcionals	9
3.3	Obstacles	10
3.4	Riscos	10
4	Metodologia i eines	11
5	Planificació	12
5.1	Recursos necessaris	12
5.2	Descripció de les tasques	13
5.2.1	Gestió del projecte	13
5.2.2	Treball previ	14
5.2.3	Disseny	14
5.2.4	Implementació	15
5.2.5	Avaluació	16
5.3	Diagrama de Gantt	17
5.4	Gestió del risc i obstacles	20
6	Pressupost	21
6.1	Costos de personal	21
6.2	Costos genèrics	22
6.2.1	Amortitzacions	22
6.2.2	Espai de treball	23
6.2.3	Total dels costos genèrics	23
6.3	Contingències	23
6.4	Imprevistos	24

6.5	Cost total del projecte	25
6.6	Control de gestió	25
7	Sostenibilitat	26
7.1	Dimensió econòmica	26
7.2	Dimensió ambiental	26
7.3	Dimensió social	27

1 Introducció i contextualització

En el moment en el qual ens trobem, on els mètodes basats en la col·lecció i interpretació de dades creixen acceleradament, un dels àmbits en què ha despertat més interès i popularitat, és en el camp de l'aprenentatge automàtic. Més en concret, en els models que tenen la capacitat de modelar dades com a grafs. Aquestes dades poden ser provinents de tota mena d'àrees del coneixement com la física, les ciències socials, la logística de transportació o les finances.

Un sistema financer pot ser definit com un sistema complex amb molts components que tenen relacions sofisticades, el qual s'adapta i reflecteix els esdeveniments que succeeixen a la realitat. Per aquestes raons, podem pensar en un mercat financer com en un graf extremadament complex, on existeixen incentius lucratius en el desenvolupament de models predictius que aconseguixin anticipar el seu comportament.

Mitjançant la tasca de classificació de nodes, el desenvolupament de models predictius amb *graph neural network* (GNN), pot ser emprat per a tota classe d'àmbits i les finances no són una excepció. Alguns exemples són la predicció d'accions en els mercats borsaris, el risc d'incompliment de préstecs o la detecció de frau financer[1]. Cal dir, que les xarxes neuronals tradicionals (*fully connected*, convolucionals CNN, recurrents RNN o LSTM) no són capaces d'aprendre d'informació que té una relació arbitrària en forma de graf, com són les finances. Per la qual cosa, les GNN tot i que són relativament noves, són aplicables a aquest problema encara que no s'ha explorat prou la utilitat i els límits.

Per totes aquestes raons, aquest treball pretén investigar el potencial i els límits de les GNN en el camp de les finances. Aquest treball de fi de grau (TFG), es situa en el marc de la *Facultat d'Informàtica de Barcelona (FIB)* de la *Universitat Politècnica de Catalunya (UPC)* i més en concret, en el grup d'investigació *Barcelona Neural Networking Center (BNN)*.

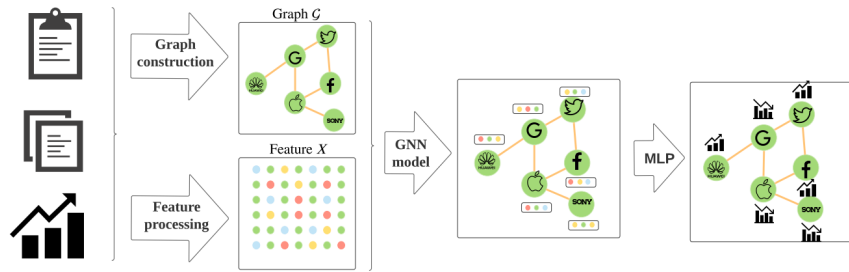


Figura 1: Desenvolupament d'una GNN per a la predicció de cotitzacions borsàries.
Font: Wang, J., Zhang, S., Xiao, Y., & Song, R. (2022). *A Review on Graph Neural Network Methods in Financial Applications* [1].

1.1 Definició de conceptes

A continuació, són definits els conceptes fonamentals per entendre el plantejament del projecte.

1.1.1 Intel·ligència artificial i aprenentatge automàtic

La intel·ligència artificial és la disciplina en el camp de les ciències de la computació que tracta de desenvolupar algoritmes i màquines que imitin la intel·ligència humana [2].

Un dels subcamps de la intel·ligència artificial és l'aprenentatge automàtic (*Machine Learning*). Aquesta branca, tracta de desenvolupar d'algoritmes estadístics que aprenguin a partir de dades i siguin capaços d'inferir i generalitzar dades no experimentades mitjançant el reconeixement de patrons.

Una de les tècniques més emprades en l'aprenentatge automàtic és l'ús de les *Graph Neural Networks* (GNN). La branca basada en les xarxes neuronals s'anomena aprenentatge profund (*Deep Learning*).

1.1.2 *Graph Neural Network* (GNN)

Les xarxes neuronals són emprades en l'àmbit de l'aprenentatge automàtic per a desenvolupar models que tinguin la capacitat de predir resultats amb un control adaptatiu. Mitjançant un conjunt de dades, els models aprenen a partir de l'experiència i infereixen conclusions a partir d'informació prèvia que pot semblar relacionada o no.

Aquestes xarxes neuronals estan inspirades en l'organització biològica de les neurones dels animals. L'homòleg de la neurona seria cada un dels nodes de la xarxa i les possibles connexions sinàptiques esdevindrien les arestes del graf.

Una *Graph Neural Network* (GNN) és un tipus de xarxa neuronal per processar dades que es poden representar com a graf.

Les GNN consten de múltiples capes, on cada una és responsable de generar una representació latent o intermèdia del node. L'intercanvi d'informació entre els nodes durant la fase d'entrenament és una de les idees fonamentals de les GNN. En cada capa es desenvolupen les següents fases:

- Transferència del missatge: cada node agrega informació provinent dels nodes veïns. El missatge estarà condicionat per les mateixes característiques del node que l'envia i dels seus nodes veïns.
- Actualització del node: cada node aprofita la informació rebuda per actualitzar la seva representació (*embedding*) a partir dels missatges rebuts dels nodes veïns.

1.2 Identificació del problema i estat de l'art

La capacitat de predir amb precisió els canvis i les tendències en finances pot ser crucial per a la presa de decisions efectiva i la gestió de riscos. Donat que els sistemes financers són altament complexos, el TFG tracta de desenvolupar models predictius mitjançant GNN.

En concret, el treball tracta d'intentar millorar els *benchmarks* actuals per a la predicció de cotitzacions borsàries. En aquest àmbit ja s'ha provat que es pot bons resultats [1] tot i que és un camp en desenvolupament. És per això que el treball tractarà de millorar el rendiment amb nous conjunts de dades mitjançant diferents tipus de GNN.

Si més no cal tenir en compte els següents aspectes clau:

- **Complexitat dels sistemes financers:** donat que existeixen molts factors que afecten el comportament dels sistemes financers cal identificar quines són les dades primordials per entrenar el model. Alguns dels reptes que es tractaran d'aconseguir són l'adaptabilitat del model a la volatilitat del mercat, la predicció de comportaments no lineals o la capacitat d'identificar patrons ocults en conjunt de dades massius.

- **Adaptabilitat de diferents GNN:** cal identificar quin tipus de GNN s'adeqüen a les característiques del sistema financer al qual es volen aplicar, ja que existeixen múltiples maneres de construir els grafs segons les particularitats de l'aplicació en qüestió (multirelacional, bipartit, dirigit, homogeni, etc.).
- **QGNN:** les *Quantum Graph Neural Network* (QGNN) són un model computacional punter que combina les característiques intrínseques de la computació quàntica i de les xarxes neuronals. Poden proporcionar certs avantatges en l'escalabilitat de les GNN en l'optimització de portfolio, el preu de derivats o la modelització de risc[3]. És un àmbit en ple desenvolupament amb el qual existeix la possibilitat d'identificar en la fase experimental certes particularitats convenients en aplicacions financeres. Si més no, els ordinadors quàntics encara no estan estandarditzats a la indústria financera perquè no són capaços de rendir millor que els ordinadors clàssics. S'estima que en els següents anys serà la primera indústria de beneficiar-se de les capacitats de la computació quàntica [3].

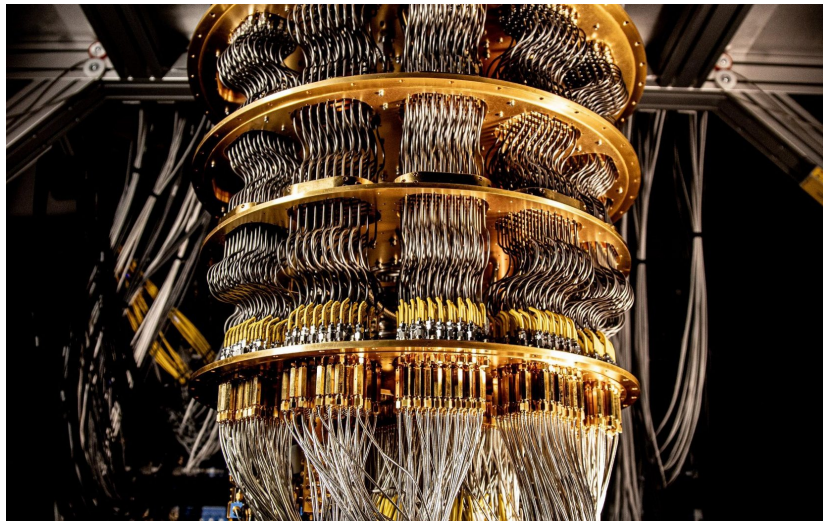


Figura 2: Fotografia d'un ordinador quàntic dissenyat per Google. Font: Titcomb, J. (2023). *The Telegraph*. [4].

- **Predicció i causalitat:** és convenient entendre que el TFG tracta de l'avaluació de models predictius. En cap cas, tracta de ser un TFG orientat a identificar causalitats, donat que aquests aspectes són tractats en el camp

de l'econometria i queda fora de l'abast del projecte.

Per a desenvolupar el projecte existeixen dues llibreries majoritàries que faciliten la programació de GNN temporals: PyTorch Geometric Temporal (PyGT) i DynaGraph. Pel que fa a la part de l'avaluació de rendiment, l'Open Graph Benchmark (OGB) proporciona conjunts de dades per a contrastar els resultats obtinguts[5]. A continuació s'esmenta un resum de la literatura existent en la predicció de cotització d'accions mitjançant GNNs que serviran per contrastar els resultats que s'obtinguin en el treball.

Característiques	Graf	Mètode	Mètrica d'avaluació
Mitjana mòbil dels últims 5/10/20/30 dies dels preus de tancament per a les empreses llistades a <i>Nikkei 225</i> [6]	Graf de relacions de proveïdors, clients, socis i accionistes	<i>Relational Stock Ranking</i>	<i>Return ratio, sharpe ratio</i>
Mitjana mòbil dels últims 5/10/20/30 dies dels preus de tancament i documents de descripció d'accions per a les empreses llistades al <i>SP500</i> i <i>NYSE</i> [7]	Xarxa de relacions basada en empreses de la Wiki	<i>Time-aware graph Relational Attention Network</i>	<i>Mean Squared Error, Mean Reciprocal Rank i Investment Return Ratio</i>
Notícies de les empreses llistades al <i>Tokyo Stock Price Index 500/100</i> [8]	Graf de correlació d'accions	<i>LSTM Relational Graph Convolutional Network</i>	Precisió
Informació de preus i xarxes socials per a les empreses llistades a l'índex <i>SP500</i> o als mercats <i>NYSE</i> o <i>NASDAQ</i> [9]	Xarxa de relacions basada en empreses de la Wiki	<i>Multipronged Attention Network for Stock Forecasting</i>	Precisió, F1, <i>Matthew's Correlation Coefficient</i>
Característiques de text i àudio de les <i>earnings call</i> per a les empreses a l'índex <i>SP500</i> [10]	Graf d' <i>earning calls</i> d'accions	<i>Volatility forecasting via Text-Audio fusion with Graph convolution networks for Earnings calls</i>	<i>Mean Squared Error, R-squared</i>
Notícies i atributs per a les etiquetes d'accions [11]	Graf de co-ocurrència de notícies	<i>Hybrid Attention Network</i>	Precisió, <i>Matthew's Correlation Coefficient</i>

Taula 1: Taula resum de la literatura existent en la predicció de cotització d'accions mitjançant GNNs [1]. Elaboració pròpia.

1.3 Agents implicats

Aquest projecte pot interessar als investigadors i les empreses que tracten d'estudiar l'aplicabilitat de l'aprenentatge automàtic i les GNN.

Per altra banda, la predicció de les cotitzacions borsàries pot interessar no només a empreses (com fons d'inversió o *trading firms*) i institucions financeres sinó, que qualsevol individual es podria beneficiar de lucrar-se amb aquests mètodes mitjançant la inversió algorítmica.

Cal esmentar que la possibilitat de trobar bons resultats en l'aplicació de *Quantum Neural Networks* també seria un avanç en el grup de recerca del N3Cat, ja que s'investiga les capacitats de l'aplicabilitat de les QGNN i de les GNN en general. Les persones directament implicades en el desenvolupament del projecte són:

- el Sergi Abadal (director del TFG)
- l'Axel Wassington (codirector del TFG)
- el Rupayan Bhattacharjee (investigador de QGNN)
- el Miquel Muñoz (autor del TFG)

2 Justificació

Les raons per les quals s'ha triat desenvolupar GNN per a la predicció de dades en sistemes financers són les següents:

- **Eficiència:** s'ha demostrat que les GNN poden aportar bones prediccions en el processament i anàlisi de dades financeres, donada la capacitat de capturar relacions no lineals i patrons ocults en conjunt de dades massius.
- **Flexibilitat:** les GNN poden ser adaptades i entrenades amb nous conjunts de dades per a abordar diferents tipus de problemes no experimentats fins al moment. Això inclou a dades de noves entitats o mercats financers. També hi ha la capacitat de diferents tipus de construcció dels grafs per abordar aquests problemes.
- **Millora de la precisió:** poden ser una eina molt rellevant per a la comparació amb sistemes tradicionals basats en l'anàlisi fonamental. És molt significatiu, en un entorn tan dinàmic i competitiu com són els mercats

financers. En concret, tot i que la literatura revisada consta de diferents mètriques d'avaluació, la precisió dels models és d'entre el 55% i el 60% [8][9][11]. Aquests resultats superen a la predicció aleatòria que té una precisió del 50%. Malgrat això, els resultats donen espai a millora atès que són nivells de precisió suficientment baixos.

- **Evolució del camp:** amb el creixent interès i avanç continu que tenen les xarxes neuronals i l'aprenentatge profund suposa una oportunitat rellevant amb un nínxol de coneixement en el qual queda molt per explorar. Cal entendre que l'evolució de l'aprenentatge automàtic i en conseqüència l'entrada de nous algoritmes i estratègies d'inversió, afecta els algoritmes existents dels altres competidors que interactuen en el mercat. D'aquesta manera algoritmes guanyadors en el passat poden quedar obsolets amb el pas del temps. Això proporciona una oportunitat contínua per tractar de batre el mercat.

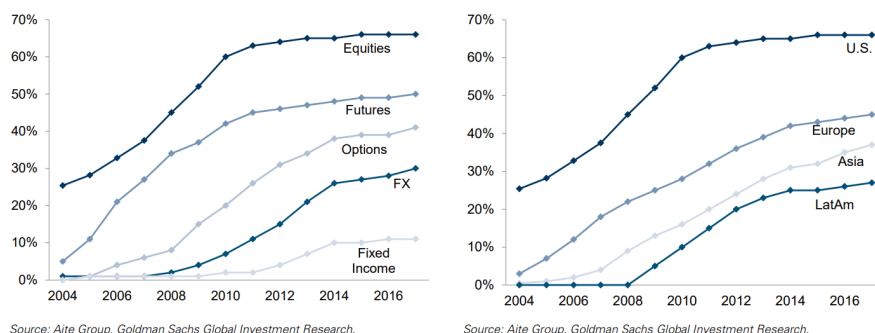


Figura 3: Percentatge d'inversió algorítmica per tipus d'actiu o de país. Font: Goldman Sachs (2017). *Top of Mind, Global Macro Research* [12].

La figura 4, és un gràfic elaborat pel grup de banca d'inversió *Goldman Sachs* on es mostra com la inversió algorítmica està en una tendència creixent en els últims anys, arribant aproximadament fins al 70% del volum de mercat global d'accions. Aquestes dades corroboren l'auge que han tingut les ciències computacionals en aquest sector.

3 Abast

3.1 Objectius i subobjectius

Els objectius del projecte són:

- Desenvolupar un programari (*pipeline*) basat en GNN per a la predicció de cotitzacions d'accions borsàries.
- Investigar l'estat de l'art actual en el camp de l'aprenentatge automàtic aplicat a la predicció de cotitzacions d'accions borsàries.
- Tractar de millorar els *benchmarks* actuals de precisió en la predicció de cotitzacions d'accions borsàries.

Com a subobjectius es poden destacar:

- Investigar i seleccionar diferents mètodes per a la construcció de grafs segons les particularitats de les dades (multirelacional, bipartit, dirigit, homogeni, etc.).
- Investigar el rendiment de diferents algoritmes de GNN: *Graph Convolutional Network (GCN)*, *Graph Attention Network (GAN)* i *Graph Isomorphism Network (GIN)*.
- Investigar possibles eficiències d'escalabilitat en utilitzar QGNNs.

3.2 Requeriments funcionals i no funcionals

A continuació es llisten els requisits funcionals i no funcionals mitjançant el model FURPS (*Functionality, Usability, Reliability, Performance, Supportability*). La primera lletra del següent llistat indica l'atribut i el número identifica l'element dins de cada categoria.

F1: Generar un conjunt de dades d'índexs borsaris com a graf a partir de URLs i arxius CSV.

F2: Que siguin capaços de ser entrenats diversos models de GNNs utilitzant el graf generat per adaptar els paràmetres.

F3: Que els models puguin ser validats per adaptar els hiperparàmetres amb dades diferents a la fase d'entrenament.

F4: Que els models siguin capaços de realitzar inferències amb nous conjunts de dades borsàries (diferents de les que s'ha entrenat i validat el model) sobre la cotització de les accions borsàries emprant el model entrenat (fase de testeig).

U1: Que el programari pugui ser controlat mitjançant una interfície de comandes intuïtiva.

U2: Oferir documentació clara i completa de com usar el programari i les seves funcionalitats.

R1: Implementar mecanismes de control i maneig d'excepcions per informar a l'usuari i evitar errors inesperats.

P1: Que la precisió obtinguda del model sigui major al 60%.

P2: Que el programari pugui ser executat en un ordinador amb 16 GB de memòria RAM.

S1: Proporcionar suport per a l'escalabilitat del sistema, permetent el processament eficient de grans conjunts de dades.

3.3 Obstacles

- **Millorar l'estat de l'art:** Tot i que s'ha comprovat que les GNN són aplicables a diversos àmbits de les finances, suposa un repte elaborar un model exitós amb nous conjunts de dades i característiques pròpies que millori l'estat de l'art.
- **Inexperiència:** la familiarització amb les GNN és un obstacle, donat que l'autor només té poca experiència amb diferents tècniques d'aprenentatge automàtic però no en GNNs.
- **Grup d'investigació i finances:** l'àmbit de les finances suposa un repte per al grup d'investigació *Barcelona Neural Networking Center (BNN)*, atès que no està especialitzat amb aquest aspecte ni hi ha una línia d'investigació específica dedicada a aquesta aplicació (tot i que l'autor també està estudiant un grau en economia).

3.4 Riscos

- **Gestió del temps:** l'autor del TFG està cursant un grau en economia a la vegada que realitza el TFG. Per aquesta raó, poden sorgir pics d'activitat inesperats que forcin canvis en la planificació del treball.
- **Complexitat de les tècniques algorítmiques:** donat que l'autor té certa inexperiència en l'àmbit de GNNs i són tècniques algorítmiques amb certa dificultat, és possible que requereixi més temps en certes tasques d'implementació del que estava previst inicialment.

- **Accidents:** encara que pugui semblar irrellevant per la baixa probabilitat que succeeixi, s’ha considerat que el risc de patir un accident seria un contratemps important que afectaria la planificació del projecte atès que el temps és molt limitat.

4 Metodologia i eines

El treball està plenament enfocat a la investigació d’aprenentatge automàtic en l’aplicació en finances. Atès que és un camp altament complex on és complicat identificar patrons, cal una metodologia molt flexible que permeti implementar algorismes, provar-los i que puguin ser ràpidament descartats en cas d’identificar mals resultats.

Per aquestes raons, se seguirà la metodologia àgil. Setmanalment, els participants del projecte es reuniran i s’analitzarà el compliment dels objectius. Depenent de les circumstàncies, amb l’ajuda dels indicadors, es faran les modificacions necessàries en la planificació del projecte o l’especificació de les tasques per poder adaptar el treball ràpidament en funció dels resultats obtinguts.

Per altra banda, com que el director i el codirector són experts en el camp de les GNNs podran ajudar ràpidament a l’autor del TFG davant de petites dificultats que puguin sorgir sense haver d’esperar a la següent reunió setmanal. Això facilitarà un accelerament en el ritme de treball i en el compliment de la planificació establerta inicialment.

A part de les reunions setmanals per fer el seguiment s’utilitzarà l’eina *Ganttter* [15] per a la planificació de les tasques. El diagrama de Gantt servirà per realitzar el control de les tasques previstes i anirà sent actualitzat dinàmicament davant dels imprevistos que sorgeixin. També s’obrirà un repositori a *GitHub* que permetrà tenir una còpia de seguretat a l’equip i també facilitarà que es pugui fer un seguiment de les actualitzacions del codi de manera eficient i ràpida. No s’ha cregut oportú fer ús d’eines més complexes com Jira perquè no s’ajusta a la casuística de l’equip. El temps del qual es disposa és molt limitat i suposa una corba d’aprenentatge molt pronunciada que podria suposar més obstacles que beneficis en l’agilització del projecte.

5 Planificació

Amb l'objectiu d'acabar el treball de fi de grau en el termini establert per la facultat i fer-ho assolint els propòsits del treball, aquesta secció tracta de definir una planificació temporal del projecte dividint-lo en tasques.

Donat que la FIB estima que la càrrega de treball són aproximadament 30 hores per crèdit i el TFG consta de 18 crèdits, el desenvolupament del projecte tindrà una durada aproximada de 540 hores. El començament del projecte (junt amb l'inici del quadrimestre de primavera) va ser el dia 12 de febrer i el dia límit per acabar-lo ha de ser el 25 juny (considerant la data de lectura més propera tot i que encara no ha estat assignada). Atès que tracta d'un projecte d'investigació dins del grup de recerca *Barcelona Neural Networking Center* (BNN), no hi ha restriccions temporals afegides de cap client o altres circumstàncies extraordinàries.

5.1 Recursos necessaris

Pel que fa als recursos necessaris per a desenvolupar el projecte, calen esmentar els següents requeriments de personal:

- El director del projecte és l'encarregat de dirigir i coordinar l'equip. També, orienta per definir l'abast del projecte i supervisa les tasques que es realitzen setmanalment per conduir el TFG davant de possibles imprevistos.
- El codirector del projecte és l'encarregat d'orientar i ajudar al programador en possibles obstacles en el qual el programador li pugui convenir certa ajuda en tasques específiques i tècniques. Junt amb el director, també supervisa i ajuda a definir l'abast del projecte.
- El programador és l'encarregat de programar les *GNN* en un sistema financer en el qual proporcionin bons resultats.

Per altra banda, cal esmentar el següent material imprescindible:

- Tres ordinadors per a cada un dels treballadors en el projecte.
- Software: Només es preveu utilitzar *software* gratuït: *Google Scholar* i *ArXiv* (per accedir a la literatura de l'estat de l'art i els algoritmes necessaris), *Kaggle* (per als *datasets*), *Visual Studio Code* (edició de codi), *GitHub* (còpia de seguretat i seguiment del codi), PyTorch Geometric

Temporal(PyGT), DynaGraph (llibries per a GNNs temporals) i *Python* (llenguatge de programació).

5.2 Descripció de les tasques

En aquest apartat, es detallen les tasques necessàries per a realitzar aquest treball de fi de grau exitosament. Cada una de les tasques és identificada amb un codi, una explicació i una estimació d'hores requerides per a complir-la.

5.2.1 Gestió del projecte

La gestió del projecte esdevé fonamental per a dur a terme totes les tasques administratives. L'assignatura obligatòria *GEP* de la *Facultat d'Informàtica de Barcelona* (FIB) forma a l'estudiant per a poder gestionar el projecte de fi de grau i també futurs projectes professionals. Sumant la dedicació de les tasques de gestió del projecte descrites a continuació, obtenim un total de 120 hores.

- **Contextualització i abast (GP1):** en aquesta tasca es defineixen i s'acoten els objectius del projecte. S'estudia la rellevància del tema en qüestió per a justificar el perquè de l'elecció. També s'indiquen els mitjants amb els quals es desenvoluparà el projecte. És necessari haver fet un estudi de l'estat de l'art per a poder avaluar la rellevància del projecte. La duració de la tasca ha estat de 30 hores.
- **Planificació (GP2):** en aquesta tasca es defineixen els recursos necessaris i la descripció de les fases i tasques del projecte. S'estudien les dependències de les tasques per a poder establir una planificació que sigui coherent amb les dates límit i requisits del projecte. També s'esmenten els riscos i obstacles per a poder prevenir-los i resoldre'ls. La duració de la tasca ha estat de 8 hores.
- **Pressupost i sostenibilitat (GP3):** s'estima el pressupost del projecte tenint en compte totes les tasques i requisits del punt anterior. A més es justifica la sostenibilitat del projecte. La duració de la tasca ha estat de 7 hores.
- **Memòria (GP4):** és una part fonamental del TFG, ja que consta dels documents característics de GEP a més de l'explicació tècnica del projecte i del que l'envolta. La dedicació és alta i és una part troncal, per tant, es preveu un total de 60 hores.

- **Presentació (GP5):** consisteix a preparar la defensa que tindrà lloc durant la lectura del TFG. Donat que només caldrà revisar el treball fet i preparar la presentació oral, s'estima un total de 5 hores de preparació.
- **Reunions (GP6):** consisteixen en trobades setmanals remotes o físiques en les quals es reuneixen el codirector, director i programador per fer un seguiment del projecte. La durada cada reunió és de mitja hora, per tant, s'estima un total de 10 hores.

5.2.2 Treball previ

Consisteix en la fase inicial del projecte. En aquest punt es tracta d'aprendre els coneixements necessaris i investigar en el projecte per a estar preparat per a desenvolupar-lo. Sumant les hores previstes per a aquesta secció, obtenim un total de 70 hores estimades.

- **Estudi de l'estat de l'art (TP1):** considerant que encara no s'ha investigat suficient en l'abast i les limitacions de les GNN tampoc en les aplicacions en finances, és necessari una fase d'investigació per saber que és el que s'ha aconseguit. D'aquesta manera es podran trobar nínxols de coneixement per determinar que pot aportar el TFG. Tenint en compte que requereix certa dedicació, s'ha estimat un total de 30 hores.
- **Familiarització amb GNN i llibreries (TP2):** tenint en compte que el programador té coneixements en aprenentatge automàtic però mai ha programat GNNs directament, suposa una tasca de preparació. També, és necessari familiaritzar-se amb *PyTorch Geometric Temporal (PyGT)* i *DynaGraph*, atès que seran llibreries necessàries per a desenvolupar el projecte. Considerant l'alta complexitat d'aquesta tasca, s'ha estimat un total de 40 hores.

5.2.3 Disseny

En la secció de disseny, el temps total estimat de les següents tasques és de 30 hores.

- **Disseny de conjunts de dades (*datasets*) (D1):** cal decidir quins mercats financers, índexs o altres dades que puguin proporcionar informació en la predicció de la cotització de les accions són adequades per a

formar part de l'*input* del algorisme. El temps esperat de la tasca (10 hores) no és molt alt perquè no requereix una alta dedicació.

- **Disseny de les GNN (D2):** cal definir l'algorisme de les GNN. Això implica decidir com es construeix el graf (multirelacional, bipartit, dirigit, homogeni, etc.) i quina modalitat de GNNs s'utilitza. Després de revisar la literatura en l'àmbit, el temps estimat és de 20 hores.

5.2.4 Implementació

La implementació consisteix a traduir el disseny de les funcionalitats a codi. Aquesta és la part més llarga i essencial del projecte. Per això, el temps total estimat és de 180 hores. Mitjançant les llibreries PyTorch Geometric Temporal (*PyGT*) i *DynaGraph*, amb el llenguatge de programació *Python* cal programar l'algorisme de les GNNs dissenyat anteriorment. Es desenvoluparan 3 tipus de GNN: *Graph Convolutional Network (GCN)*, *Graph Attention Network (GAN)* i *Graph Isomorphism Network (GIN)*.

- **Implementació dels conjunts de dades (*datasets*) (I1):** cal trobar les dades o formatejar-les manualment perquè estiguin en un format llegible per l'algorisme (per exemple *csv* o *json*).
- **Implementació de la *Graph Convolutional Network (GCN)* (I2):** aquest tipus de GNN extreu informació de la matriu d'adjacències del graf per extreure característiques de la relació dels nodes (propagant la informació a través del graf). Com que és la primera GNN que es desenvoluparà té una càrrega lleugerament més elevada d'hores (60 hores).
- **Implementació de la *Graph Attention Network (GAN)* (I3):** aquest tipus de GNN permetrà assignar pesos diferents als nodes per poder donar rellevàncies desiguals als nodes. Potencialment, podria ajudar a millorar els resultats obtinguts a I2. S'ha estimat un total de 50 hores per acomplir la tasca.
- **Implementació de la *Graph Isomorphism Network (GIN)* (I4):** amb aquesta variant es contribueix a determinar la similitud estructural entre grafs. En aquest cas, podria ajudar a identificar tendències del mercat. Com a I3, s'ha estimat un total de 50 hores per a dur a terme la tasca.

5.2.5 Avaluació

- **Mètrica d'avaluació (A1):** cal justificar quina mètrica d'avaluació és rellevant per a l'estudi i implementar-la per obtenir resultats (*Mean Squared Error*, *return ratio*, precisió, etc.). La dedicació estimada és de 20 hores.
- **Comparació i millora de resultats GCN (A2), GAN (A3) i GIN (A4):** cal comparar els resultats aconseguits amb els que pertanyen a l'estat de l'art. En cas de no superar-los en cap aspecte, tractar de trobar alguna manera de millorar-los. Aquesta tasca serà realitzada per igual a cadascuna de les GNN descrites anteriorment: GCN (A2), GAN (A3) i GIN (A4). Cal dividir-la en tres parts, tot i que el procediment sigui el mateix a les 3, atès que ens interessa anar tancant parts del treball com més aviat millor. D'aquesta manera serà possible planificar i mitigar els riscos adequadament. La dedicació d'hores serà de 40, 30 i 30 hores respectivament. Això és degut al fet que el primer cop que es fa la tasca tindrà una càrrega d'hores més elevada.
- **Validació (A5):** cal fer un estudi exhaustiu de la correctesa dels algorismes per a detectar possibles errors en el treball. Aquesta tasca es realitza paral·lelament amb la implementació per a poder rectificar a temps. La dedicació estimada és de 20 hores.

Codi	Tasca	Temps(h)	Dependència
GP	Gestió del projecte	120	
GP1	Contextualització i abast	30	
GP2	Planificació	8	GP1
GP3	Pressupost i sostenibilitat	7	GP1, GP2
GP4	Mèmorria	60	
GP5	Presentació	5	
GP6	Reunions	10	
TP	Treball previ	70	
TP1	Estudi de l'estat de l'art	30	
TP2	Familiarització amb GNN i llibreries	40	
D	Disseny	30	
D1	Disseny de conjunts de dades (<i>datasets</i>)	10	TP1
D2	Disseny de les GNN	20	TP1
I	Implementació	180	
I1	Implementació dels conjunts de dades (<i>datasets</i>)	20	D1
I2	Implementació de la GCN	60	TP2, D2, I1
I3	Implementació de la GAN	50	TP2, D2, I1
I4	Implementació de la GIN	50	TP2, D2, I1
A	Avaluació	140	
A1	Mètrica d'avaluació	20	D2
A2	Comparació i millora dels resultats (GCN)	40	I2
A3	Comparació i millora dels resultats (GAN)	30	I3
A4	Comparació i millora dels resultats (GIN)	30	I4
A5	Validació	20	TP2, D2, I1

Taula 2: Taula resum de la planificació amb les dependències corresponents. Elaboració pròpia.

5.3 Diagrama de Gantt

Per desenvolupar el diagrama s'ha tractat de preveure una dedicació de mitjana de 6-7 hores al dia. D'aquesta manera, la implementació i avaluació del projecte acaba 3 setmanes abans del projecte. Tal com es comentarà posteriorment, existeixen probabilitats significatives que el projecte s'hagi d'allargar en quantitat

d'hores i així es mitiguen aquests riscos. També es deixa una setmana sencera per revisar la memòria a principis de juny.

Tot i que no hi ha dependència entre la implementació de les 3 alternatives de GNN, s'implementen seqüencialment donat que així es pot anar fent la comparació i millora de resultats mentre s'implementa la següent GNN. Amb aquesta estratègia, cada GNN té una data prevista de finalització diferent que permet distribuir les tasques per poder prevenir riscos. Seria una mala estratègia desenvolupar les 3 GNN alhora perquè potser, en comparar als resultats al final de tot, ja no quedaria temps per acabar el projecte a la data prevista.

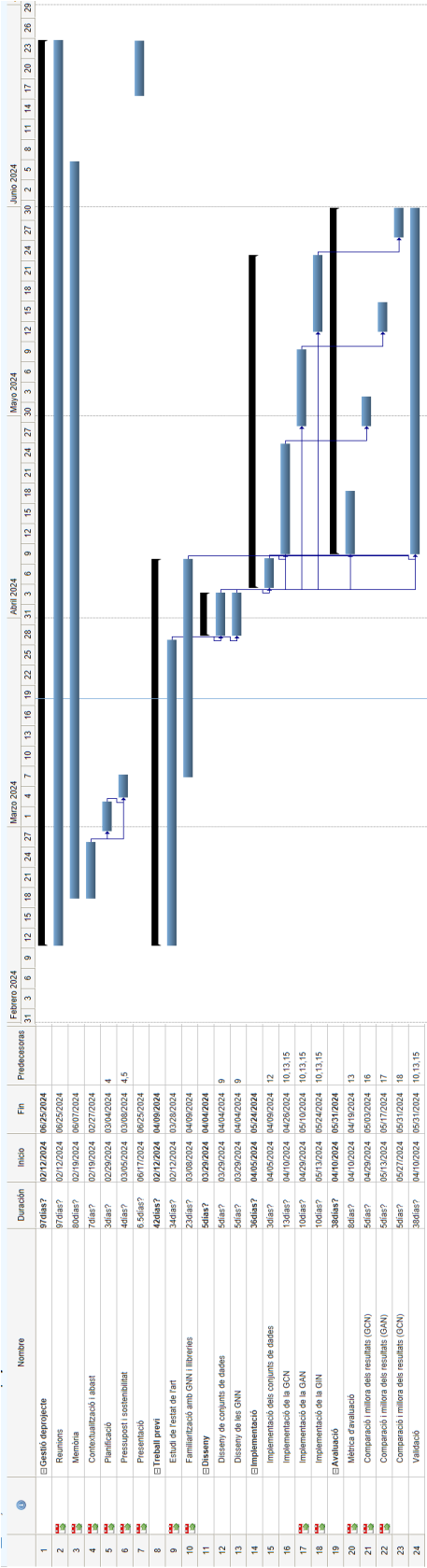


Figura 4: Diagrama de Gantt elaborat mitjançant *Gantt*. Elaboració pròpia.

5.4 Gestió del risc i obstacles

Recuperant els riscos especificats al punt 3.4, els 3 riscos principals són: la gestió del temps, la complexitat de les tècniques algorítmiques i els accidents.

- **Gestió del temps:** els punts més àlgids d'activitat de feina al grau d'economia (UPF) són entre el 10 i 20 de juny pels exàmens finals. És una de les raons per les quals s'ha intentat planificar la finalització del projecte quan més aviat millor. Qualsevol altre imprevist en unes altres dates en aquest aspecte, simplement allargaria la planificació d'aquella setmana o aquells dies, posposant totes les posteriors tasques. No afectaria els terminis pels marges que s'han planificat perquè no allargarien el projecte més d'una setmana.
- **Complexitat de les tècniques algorítmiques:** això podria suposar un augment d'hores en la implementació o el disseny de les GNNs. L'estratègia utilitzada per implementar les 3 GNNs ajuda a mitigar aquest risc. Tal com s'ha esmentat en el punt anterior, la concurrència d'implementar la següent GNN mentre es milloren els resultats de la GNN anterior permet ajustar la planificació en cas que alguna d'aquestes tasques s'allargui. Si s'allargués el disseny o la implementació, sí que afectaria les tasques que en són dependents endarrerint-les (no hi hauria problema perquè el marge previst s'ha ajustat per preveure aquests riscos). S'ha estimat un possible increment de 10 hores en disseny i un potencial augment de 30 hores d'implementació que allargarien la finalització de la implementació, comparació i validació de resultats una setmana en el pitjor dels casos. Per una anàlisi de les probabilitats dels esdeveniments descrits i com afecten el pressupost cal veure el punt 6.4.

En la familiarització amb GNN i llibreries o l'estat de l'art, no s'espera un increment d'hores en cap cas, perquè s'ha deixat molt espai temporal per madurar aquestes tasques. En cas que els resultats d'alguna de les 3 GNNs no superessin a l'estat de l'art, no comportaria un augment de la tasca de comparació de resultats, ja que cap la possibilitat de que els *benchmarks* de l'estat de l'art no siguin millorables per alguna de les GNNs proposades.

- **Accident:** a efectes pràctics un accident endarreriria el projecte de la mateixa manera que el punt de *Gestió del temps* en el projecte. Òbviament,

es complicat estimar la quantitat de temps que obstruïria la possibilitat de continuar endavant amb el treball, però el marge que s’ha deixat al final de la planificació també mitigaria aquest risc en certa manera encara que la probabilitat que succeeixi sigui baixa.

6 Pressupost

6.1 Costos de personal

Tal com s’ha esmentat anteriorment, es durà a terme el treball amb un equip format pel director, codirector i l’autor del TFG. Tot i això, per a calcular el pressupost, cal diferenciar entre tres rols diferents: el cap de projecte, el programador i el director de recerca.

El programador i el cap de projecte en aquest cas serà exclusivament l’autor del TFG. No obstant, la feina del director de recerca serà compartida pel director i el codirector. S’ha calculat els salaris que es poden observar a la taula 3 segons les dades de *Talent.com*. També s’ha calculat el cost de la seguretat social (SS) que és un 30% afegit.

Rol	Salari brut	SS	Retribució
Programador	14,62€	4,39€	19,01€
Cap de projecte	19,49€	5,85€	25,34€
Director de recerca	20,29€	6,09 €	26,38€

Taula 3: Salari de cada rol segons dades de *Talent.com*. Elaboració pròpia.

Després d’establir el cost per hora de cada rol, podem calcular el cost per activitat (CPA). A la taula 4 es pot observar el cost de cadascuna de les tasques planificades al diagrama de Gantt anteriorment segons el cost de cada rol per dur-la a terme. La implementació de GNN i la comparació i millora de resultats s’ha definit de forma compacta (sense separar per tipus de GNN), ja que no era necessari desglossar les tasques com s’ha fet prèviament al diagrama de Gantt.

Codi	Tasca	Cap de projecte(h)	Programador(h)	Director(h)	Cost(€)
GP	Gestió del projecte	110	10	20	3505,1
GP1	Contextualització i abast	25		5	765,4
GP2	Planificació	8			202,72
GP3	Pressupost i sostenibilitat	7			177,38
GP4	Mèmorria	55		5	1525,6
GP5	Presentació	5			126,7
GP6	Reunions	10	10	10	707,3
TP	Treball previ	10	60		1394
TP1	Estudi de l'estat de l'art	5	20		506,9
TP2	Familiarització amb GNN i llibreries	5	40		887,1
D	Disseny	30	0		760,2
D1	Conjunts de dades (<i>datasets</i>)	10			253,4
D2	GNN	20			506,8
I	Implementació	0	180		3421,8
I1	Conjunts de dades (<i>datasets</i>)		20		380,2
I2	GNN		160		3041,6
A	Avaluació	35	95	10	2956,65
A1	Mètrica d'avaluació	20			506,8
A2	Comparació i millora dels resultats	10	90		1964,3
A3	Validació	5	5	10	485,55
Total CPA		185	345	30	12037,75

Taula 4: Taula resum del cost de les tasques segons el rol. Elaboració pròpia.

6.2 Costos genèrics

6.2.1 Amortitzacions

Per desenvolupar el projecte es necessitaran 3 ordinadors portàtils per a cada un dels treballadors amb un cost estimat de 700 euros per ordinador (assumint una gamma mitjana). A part, la mitjana del temps de vida útil dels ordinadors és d'aproximadament 5 anys i les hores d'ús han estat calculades a la taula 4 (s'ha estimat una dedicació d'hores equivalent entre el director i el codirector). Tenint en compte que un any té 220 dies feiners de jornades laborals de 8 hores el $(\text{cost}/\text{hora}) = \text{cost}/(\text{vida_util} * 220 * 8)$. L'amortització d'aquest material tenint

en compte cadascun dels rols és:

Hardware	Vida útil (anys)	Temps d'ús (hores)	Amortització(€)
Ordinador autor TFG	5	530	42,16
Ordinador director	5	15	1,19
Ordinador codirector	5	15	1,19
Total			44,54

Taula 5: Amortitzacions del *hardware*. Elaboració pròpia.

6.2.2 Espai de treball

El projecte es durà a terme remotament. Tot i això, per a comptabilitzar el cost de l'espai de treball s'ha calculat tenint en compte un *coworking* (inclou internet, aigua i electricitat) amb preus de l'empresa *Aticco* (preus estàndard a Barcelona). El preu per una persona a dedicació *full time* és de 220€/mes. Per altra banda, el cost de treballar-hi un dia és de 35€. També, es calcula que l'autor del TFG treballarà durant 5 mesos una dedicació aproximada *fulltime* i que el director i el codirector treballaran un total de 2 dies laborals cadascun (15 hores de feina comptabilitzades cadascun). El **preu total** és de $220*5+2*2*35 = 1240$ €.

6.2.3 Total dels costos genèrics

La taula 6 mostra el total dels costos genèrics esmentats anteriorment.

Concepte	Cost(€)
Amortitzacions	44,54
Espai de treball	1240
Total	1284,54

Taula 6: Taula resum costos genèrics. Elaboració pròpia.

6.3 Contingències

És necessari tenir previst i calculat el fet que apareguin complicacions durant el projecte que puguin suposar un augment del pressupost. S'ha escollit un percentatge del 15 % per a preparar una partida de contingència, donat que hi

ha una probabilitat considerable de que el cost per activitat incrementi i s'hagi d'utilitzar aquesta partida de contingència (explicat al punt següent).

Tenint en compte que el total dels costos per activitat és de 12037,7€ i el total dels costos genèrics és de 1284,54€, el cost de contingència total és de 1998,34€.

6.4 Imprevistos

En aquest apartat es calcula el cost dels possibles imprevistos tenint en compte la probabilitat existent de que succeeixin. Considerant els punts de l'apartat 5.4, els apartats de *Gestió de temps* i d'*Accident* no comportarien costos addicionals al projecte.

Pel que fa a l'increment del temps del disseny, s'ha estimat un possible augment de 10 hores amb una probabilitat del 30%, ja que és significativament probable que les GNNs dissenyades no s'adeqüin a les particularitats dels mercats financers (probablement s'hauran de dissenyar múltiples models).

Per altra banda, qualsevol canvi de disseny comporta un canvi en la implementació. És significativament probable (40%) que es necessitin més hores de les calculades (un augment de 30 hores) perquè s'utilitzaran tècniques algorítmiques molt avançades no provades anteriorment per l'autor a nivell pràctic (GNNs temporals).

Per últim, cal dir que la probabilitat del fet que es necessiti un nou ordinador és molt baixa (5%) atès que no s'espera una alta demanda computacional per a executar els algorismes (també s'ha considerat la probabilitat d'un error del fabricant).

Imprevist	Cost(€)	Probabilitat	Cost esperat(€)
Nou ordinador	700*3	5%	105
Temps de disseny (+10h)	253,4	30%	76,02
Temps d'implementació (+30h)	570,3	40%	228,12
Total			409,14

Taula 7: Taula resum dels costos causats per possibles imprevistos. Elaboració pròpia.

6.5 Cost total del projecte

A continuació es calcula el cost total del projecte a la taula 8 tenint en compte tots els costos calculats anteriorment.

Concepte	Cost(€)
CPA	12037,75
CG	1284,54
Contingències	1998,34
Imprevistos	409,14
Total	15729,72

Taula 8: Taula resum cost total del projecte. Elaboració pròpia.

6.6 Control de gestió

Per realitzar adequadament el control de gestió s'aprofitaran les reunions setmanals per abastar aquest aspecte. D'aquesta manera es podrà fer un seguiment del projecte per poder corregir a temps qualsevol desviació que sigui detectada (econòmica o temporal).

S'utilitzaran els següents indicadors per fer revisions de costos. Caldrà assegurar que la desviació sigui la menor possible:

- $\text{Desviació_cost} = (\text{cost_estimat} - \text{cost_real}) * \text{hores_reals}$
- $\text{Desviació_hores} = (\text{hores_estimades} - \text{hores_reals}) * \text{cost_estimat}$

Cada setmana es farà una estimació de les hores reals dedicades en comparació amb les hores estimades a la planificació inicial (mitjançant l'eina *Ganttter*), així com els costos econòmics reals i els estimats. En cas que hi hagi desviacions desfavorables, es reajustarà la planificació degudament o s'usarà la partida de contingència per cobrir els costos addicionals.

Aleshores, si es detecta que el temps o capital restant no és suficient, es reajustarà l'abast del projecte per a complir amb els requisits monetaris i temporals.

Malgrat això, en cas que la desviació sigui favorable i la feina comporti menys hores que les estimades inicialment, s'optarà per avançar feina posterior per reduir possibles riscos i poder realitzar ampliacions en el projecte. Alguna d'aquestes possibles ampliacions, podria ser avaluar el rendiment de les *Quantum Graph Neural Networks (QGNN)*.

7 Sostenibilitat

Considero que sempre he tingut un interès i preocupació en les implicacions que poden tenir els projectes en la societat. Es pot veure reflectit per la tria d'estudiar també un grau en economia, donat que constantment s'estudia el paper de les externalitats quan s'acompleix activitat econòmica. Crec que és un aspecte al qual cada cop se li dona més importància però no la suficient. El govern i les institucions haurien de ser els encarregats de dissenyar incentius que fonamentin la sostenibilitat social i ambiental perquè no sigui només la dimensió econòmica la prioritària.

Malgrat que penso que estic altament conscienciat de la importància de la sostenibilitat en l'àmbit teòric, crec que em falta posar els meus coneixements en pràctica per poder participar en algun projecte on la magnitud de l'impacte sigui significativa. L'enquesta m'ha ajudat a entendre quin és el meu grau de coneixement de l'impacte de les TIC a la societat i a identificar les febleses que puc millorar en l'àmbit pràctic.

7.1 Dimensió econòmica

El cost del projecte ha estat desglossat detalladament anteriorment. Aleshores, s'ha pogut comprovar com el cost del projecte és adequat, ja que no es necessita *hardware* ni *software* d'alt cost. La part que destaca més és la del sou dels treballadors, per tant, no es pot reduir tenint en compte el mercat laboral actual. Per altra banda, els costos de desenvolupament (ordinadors i internet) són totalment essencials i estàndards.

No obstant, la possibilitat de predir les cotitzacions d'accions al mercat borsari podria aportar beneficis que justifiquen els costos del projecte. Seria una eina útil per a qualsevol institució, empresa o particular que participi en intercanvis borsaris.

7.2 Dimensió ambiental

El desenvolupament del projecte no requereix de cap *hardware* específic que sigui altament perjudicial per al medi ambient. Només es pot esmentar l'impacte de la fabricació dels ordinadors portàtils o el cost energètic de l'electricitat que requereixen els ordinadors.

El fet de tenir 3 ordinadors portàtils diferents s'ha considerat necessari per a poder paral·lelitzar la feina i permetre als treballadors treballar remotament

d'una manera eficient.

7.3 Dimensió social

Els impactes negatius socials poden ser la part més controvertida d'aquests 3 apartats. Tot i que ha quedat demostrat anteriorment que el *trading* algorítmic està en una tendència creixent (70% del volum global d'accions), suposa un risc per als sistemes financers i un repte per a les entitats que el regulen. Com a exemple, es pot destacar el *Flash Crash* de maig del 2010[14].

El *Flash Crash* va consistir en una caiguda del 6% dels principals índexs borsaris d'Estats Units en tan sols uns minuts. El *Dow Jones* va caure 1000 punts, constituint la major caiguda mai registrada fins aleshores. Va ser demostrat més tard que va ser causa d'una tàctica d'inversió anomenada *spoofing*. Aquesta consisteix en programar un alt volum d'ordres falses al mercat que són cancelades abans que siguin complertes. D'aquesta manera, es va alterar el volum i la percepció dels altres inversors que participen en el mercat. El *trading* algorítmic podria ser usat per magnificar aquesta mena de tècniques. També poden tenir altíssimes pèrdues en tan sols pocs segons, ja que operen amb altes quantitats de capital en períodes de nanosegons.

Tot i això, cada cop són més els controls que hi ha per part de les institucions que regulen aquests mercats per prevenir les conseqüències d'aquestes tècniques d'inversió. Tots els progressos tecnològics poden ser utilitzats amb mala fe si les entitats encarregades de regular-ho no es modernitzen amb la mateixa velocitat que el progrés científic. Considerant que el TFG està en un marc purament acadèmic en el que ni tan sols s'interactuarà amb el mercat, no comporta riscos directes per a la societat.

Per últim, també suposa un avanç que busca aprofundir en les aplicacions de l'aprenentatge automàtic. Per la qual cosa, suposaria un benefici social per a totes les institucions de recerca que treballin en aquest àmbit.

Bibliografia

- [1] Wang, J., Zhang, S., Xiao, Y., & Song, R. (2022). *A Review on Graph Neural Network Methods in Financial Applications* (arXiv:2111.15367). arXiv. <http://arxiv.org/abs/2111.15367>
- [2] John McCarthy (1956), The Dartmouth Conference.
- [3] Herman, D., Googin, C., Liu, X., Galda, A., Safro, I., Sun, Y., Pistoia, M., & Alexeev, Y. (2022). *A Survey of Quantum Computing for Finance* (arXiv:2201.02773). arXiv. <http://arxiv.org/abs/2201.02773>
- [4] Titcomb, J. (2023). *Supercomputer makes calculations in blink of an eye that take rivals 47 years. The Telegraph.* <https://www.telegraph.co.uk/business/2023/07/02/google-quantum-computer-breakthrough-instant-calculations/>
- [5] Longa, A., Lachi, V., Santin, G., Bianchini, M., Lepri, B., Lio, P., Scarselli, F., & Passerini, A. (2023). *Graph Neural Networks for temporal graphs: State of the art, open challenges, and opportunities* (arXiv:2302.01018). arXiv. <https://doi.org/10.48550/arXiv.2302.01018>
- [6] Matsunaga, D., Suzumura, T., & Takahashi, T. (2019). *Exploring Graph Neural Networks for Stock Market Predictions with Rolling Window Analysis* (arXiv:1909.10660). arXiv. <https://doi.org/10.48550/arXiv.1909.10660>
- [7] Ying, X., Xu, C., Gao, J., Wang, J., & Li, Z. (2020). *Time-aware Graph Relational Attention Network for Stock Recommendation. Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2281–2284. <https://doi.org/10.1145/3340531.3412160>
- [8] Li, W., Bao, R., Harimoto, K., Chen, D., Xu, J., & Su, Q. (2020). *Modeling the Stock Relation with Graph Network for Overnight Stock Movement Prediction. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 4541–4547. <https://doi.org/10.24963/ijcai.2020/626>
- [9] Sawhney, R., Agarwal, S., Wadhwa, A., & Shah, R. R. (2020). *Deep Attentive Learning for Stock Movement Prediction From Social Media Text and Company Correlations.* In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 8415–8426). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.676>

- [10] Sawhney, R., Khanna, P., Aggarwal, A., Jain, T., Mathur, P., & Shah, R. R. (2020). *VolTAGE: Volatility Forecasting via Text Audio Fusion with Graph Convolution Networks for Earnings Calls*. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 8001–8013). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.643>
- [11] Liou, Y.-T., Chen, C.-C., Tang, T.-H., Huang, H.-H., & Chen, H.-H. (2021). *FinSense: An Assistant System for Financial Journalists and Investors*. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 882–885. <https://doi.org/10.1145/3437963.3441704>
- [12] Goldman Sachs (2017). *Top of Mind, Global Macro Research* https://www.gsam.com/content/dam/gsam/pdfs/sg/en/commentary/GS_Top%20of%20Mind_June.pdf?sa=n&rd=n
- [13] *Salario para Programador en España—Salario Medio. (n.d.). Talent.com.* Retrieved 11 March 2024, from <https://es.talent.com/salary>
- [14] *4 Big Risks of Algorithmic High-Frequency Trading. (n.d.). Investopedia.* Retrieved 11 March 2024, from <https://www.investopedia.com/articles/markets/012716/four-big-risks-algorithmic-highfrequency-trading.asp>
- [15] *Ganttter. (n.d.).* Retrieved 10 March 2024, from <https://www.ganttter.com/>